PERSPECTIVE

Multi-scale spatial ecology analyses: a Kullback information approach

Gintarė Huckeba (ID) · **Bjarne Andresen** (ID) ·
Ty N. F. Roach

© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Context   The way organisms are patterned in space dictates the outcome of many ecological processes such as growth, survival, colonization, and migration. The field of landscape ecology has developed quantitative metrics to describe spatial patterning using the concept of entropy. However, a general theory of how these patterns relate to one another within and between different organizational levels and over different spatial scales has remained incomplete.

Objectives   Review how statistical versions of entropy have been applied to detect spatial organization and propose a theoretical framework to use Kullback–Leibler relative entropy for cross-scale analyses on a landscape of any size.

Methods   Examine previous efforts using entropy in landscape ecology and introduce a Kullback Information Index as a next step in the science of scaling.

Results   Entropic indices can provide compositional and configurational information about a system and can be used to detect landscape patterns. Yet, most entropy-based metrics are scale-dependent, highlighting the need to find a common currency for comparative analysis across scales. The non-symmetric unitless property of the Kullback–Leibler relative entropy may remedy that since it is theoretically capable of comparing variables and scales. The proposed framework can be extended to describe any system that contains scalable modules of interest, which will advance scaling in landscape ecology and other disciplines.

Conclusions   The Kullback Information Index describes landscapes' compositional and configurational patterns across scales. Since relative entropy is connected to information theory and thermodynamics, the framework's results can be interpreted within a broader ecological context.

Keywords   Spatial organization · Landscape patterns · Diversity · Information theory · Thermodynamics · Entropy

G. Huckeba
Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam, The Netherlands

B. Andresen
Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

T. N. F. Roach (✉)
Hawaiʻi Institute of Marine Biology, University of Hawaiʻi at Mānoa, Honolulu, USA
e-mail: smokinroachjr@gmail.com

T. N. F. Roach
The Viral Information Institute, San Diego State University, San Diego, CA, USA

Introduction

From cells to ecosystems, spatial structure plays a part in mediating the flow of energy and information (James and Fortin 2013). The organizational patterns—the composition and configuration of biological entities—often influence interaction outcomes, which can be readily described using statistical measures such as entropy (Ulanowicz et al. 2006; Harte 2011; O'Connor et al. 2019). The term entropy originated in thermodynamics aiming to describe energy transformation in an isolated system (Jaynes 1980; Lesne 2014). Boltzmann further extended thermodynamic entropy to statistical mechanics, introducing macrostate and microstate concepts that define a system's conditions at a macroscopic scale (a macrostate) and a more detailed depiction of the system's conditions at the microscopic scale (a microstate), resulting in the following equation for Boltzmann entropy:

$$S = k \log W \tag{1}$$

where k defines the units for S. In the international SI system, S has units J/K in which case k is Boltzmann's constant ($1.38\ 9 \times 10^{-23}$ J/K), and W is the number of possible microstates corresponding to the particular macrostate (Boltzmann 1872).

Almost 100 years later, a parallel statistical version of entropy was developed by Claude Shannon (1948). Outside its original scope, Shannon entropy (sometimes referred to as information entropy) was quickly adopted by many scientific disciplines, including biology (Sherwin 2010; Chanda et al. 2020; Roach 2020). It became popular as a diversity measure because a single equation could describe the distribution of different types of organisms in a community (Jost 2006; Daly et al. 2018). Shannon's formulation calculates the sum of the probabilities (p_i) that an organism belongs to a specific type (the ith species) multiplied by the log of that probability. The equation then calculates a summation over all types of organisms and represents the average uncertainty per microstate (h) as:

$$h = -\sum_{i=1}^{n} p_i \log p_i \tag{2}$$

When multiplied by the number of microstates (N), this yields Shannon entropy (H). Thus,

$$H = Nh \tag{3}$$

The field of landscape ecology has integrated entropy approaches to detect spatial patterning, unpredictability, and pattern-scale dependence (Vranken et al. 2015). However, the use of entropy in landscape ecology has been variable with some researchers arguing that Shannon entropy is 'aspatial' (sensu Leibovici 2009) as it only captures the probability distribution of the components within a system (Claramunt 2005; Wang and Zhao 2018). Recent publications advocate for Boltzmann entropy being used instead of Shannon entropy for spatial analysis, as Boltzmann entropy provides compositional as well as configurational information with an overall stronger thermodynamic interpretation (Vranken et al. 2015; Gao and Li 2019). However, Boltzmann and Shannon entropies have been derived, and proven equivalent, for general physical situations (Jaynes 1957; McQuarrie 2000; see Online Appendix A).

Regardless of which entropy equation is used, both Shannon and Boltzmann entropies fundamentally measure uncertainty, and provide information about the system. However, both entropies assume that the system is uniform or take the entropies to be averages over the system considered, thus limiting calculations to a single scale. The results then can be difficult to compare across different levels of hierarchical organization or scale (Rajaram et al. 2017). Since many ecological processes operate on multiple scales, creating different patterns at individual organizational levels, there is a need to address the multi-scale nature of ecosystems (Wu et al. 2000; Roach et al. 2018). The ability to use entropy statistics to detect spatial patterns on different scales would greatly advance the understanding of pattern-process relationships. Here, we evaluate how statistical variants of entropy have been applied in previous studies to capture spatial variability. We subsequently build on this and develop a theoretical framework to describe configurations across multiple spatial and organizational scales using a *Kullback Information Index*.

Inspired by Boltzmann entropy

In statistical mechanics, Boltzmann entropy represents an ensemble quantity that describes the variety of particles within a system. In biology, this can

also be interpreted as the number of ways constituents fill functional or physical space i.e., the degree of spatial heterogeneity of the biological entities in a system (Roach et al. 2017, 2019). In landscape ecology, Cushman (2016) found that spatial organization can be empirically captured by modifying the original Boltzmann entropy equation into

$$H_C = \ln(W) \qquad (4)$$

where H_C is calculated as a natural logarithm of the number of unique arrangements within a landscape (W, or microstates) that produce the same total edge (TE) length (a macrostate).

Cushman's formalism of Boltzmann entropy is based on the following assumptions:

Assumption 1: when entities of the same class are maximally aggregated, TE between different entities is low and therefore entropy is low.
Assumption 2: when different entities are maximally dispersed, TE increases with fewer ways to arrange entities to retrieve maximal edge length, therefore the landscape is said to have low entropy.
Assumption 3: when the position of entities is spatially random, intermediate levels of aggregation occur, creating higher entropy.

The H_C entropy calculations in Table 1 and Fig. 1 demonstrate that the maximally aggregated landscape A showed lower entropy values than landscapes C and D, however the difference between A and C was only minor. The entropy of maximally heterogeneous landscape B was 0 since landscape B had only one way to attain the maximum TE of 12. Finally, landscape D had the highest entropy, because in spatially random configurations there is a higher chance of cells of the same type to be more adjacent to each other, thus, lower TE and higher entropy.

The introduction of Boltzmann entropy to landscape ecology (Cushman 2016) provided new methods to investigate landscape structuring using thermodynamic principles. Initially this method (Cushman 2016) was limited to small landscapes, as possible arrangements dramatically increased at larger scales. Further work (Cushman 2018) proposed an alternative method of computing Boltzmann entropy by randomizing landscape mosaics, counting the proportion of landscape microstates, and finally calculating relative (Boltzmann) entropy. Recent reviews (Gao and Li 2019; Gao et al. 2021) agree that Cushman's (2018) method is effective but still lacks experimental evaluation.

Efforts to extend Cushman's entropic index for landscapes represented by gradients were made by Gao et al. (2017). Here, a macrostate was defined as the upscaled depiction of a landscape, and

Table 1 Entropy results with changing resolution of Fig. 1 landscapes

Scale	Landscape	TE	H	H_C	H_{Cl}	H_{WZ}
3×3	A	4	0.991	2.48	0.754	2.786
	B	12	0.991	0	1.169	118.929
	C	8	0.991	2.56	1.079	12.459
	D	7	0.991	3.58	1.053	15.417
4×4	E	4	1	–	0.676	2
	F	24	1	–	1.108	240
	G	12	1	–	1.039	21.466
	H	10	1	–	1.005	17.889
8×8	I	8	1	–	0.641	2
	J	112	1	–	1.029	1120
	K	23	1	–	0.951	17.011
	L	20	1	–	0.934	17.889
16×16	M	16	1	–	0.632	2
	N	480	1	–	1.007	4800
	O	46	1	–	0.933	17.011
	P	48	1	–	0.917	21.466

Here, H refers to Shannon entropy, H_C proposed by Cushman (2016), H_{Cl} represents Claramunt (2005) spatial diversity and H_{WZ} depicts spatial entropy as proposed by Wang and Zhao (2018). Calculation performed with Rstudio (see Supplementary material)
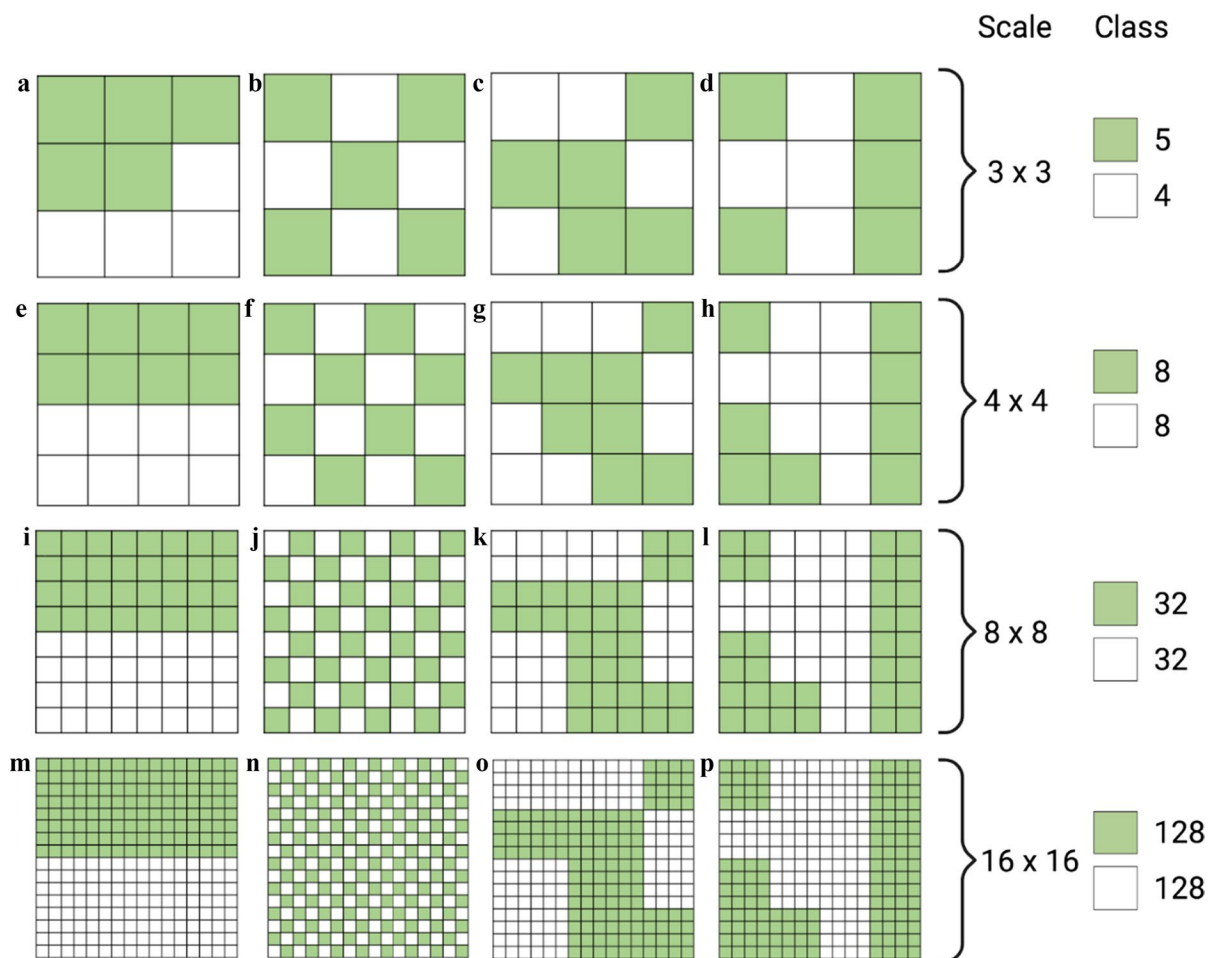
Fig. 1 Sixteen chessboard-like landscapes **a–p** with dimensionality of 3×3, 4×4, 8×8 and 16×16 with two classes –class white, and class green. The existing 3×3 landscapes

a–d were fine-grained to 4×4, 8×8, and 16×16 to depict how entropy changes when changing spatial resolution while keeping the configurations similar to the original **a–d** landscapes

microstates were calculated by finding all possible ways to downscale to the original macrostate. Gao et al.'s (2017) treatment of Boltzmann entropy allowed for a more general theory as it did not depend on how one partitions the landscape. For example, calculating the distribution among some larger groups and subsequently the micro distribution within each group should yield the same results as estimating the full probability distribution of a single large set. In general, Cushman's and Gao's work made it feasible to calculate the Boltzmann entropy of landscape mosaics and gradients. These computational methods were later reviewed by Gao and Li (2019) concluding that future research should focus on a general model-independent

calculation of Boltzmann entropy with a deeper thermodynamic interpretation.

Overall, Boltzmann entropy's link to thermodynamics makes the equation attractive for broader applications. However, Boltzmann entropy becomes computationally intensive when applied to realistic landscapes because the resolution and number of classes are often higher than in theoretical experiments (Zhang and Wu 2020). Nevertheless, steps have been taken to optimize computation times of Boltzmann entropy with proposal of the R package '*belg*' and use of the Wasserstein metric (Nowosad et al. 2020; Gao et al. 2021). Moreover, landscapes that have different TE values may still have the same number of unique configurations, which leads to equal

entropy values [see in Cushman (2016) Table 1]. Instead of using TE as the main spatial factor, a few different approaches have been proposed using spatial entropy, a term used to describe an entropic variant that is capable of capturing landscape heterogeneity using other spatial variables.

Spatial entropy

Although the use of Shannon entropy for spatial analysis has been questioned, it is important to note that statistically the Shannon entropy equation is valid in its simplest form and can still be useful (Zhao and Zhang 2019). A few different approaches have been proposed to improve Shannon entropy by incorporating spatial features to elucidate how different components of the sample are spatially organized (Wang and Zhao 2018). For instance, Claramunt (2005) used distance as a factor to determine the degree of similarity between entities in space. Claramunt based this on Tobler's "First Law of Geography" (TFL), which states that "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970). Accordingly, a spatial measure of entropy was calculated by following two assumptions:

Assumption 1: when different entities are closer, diversity (i.e., entropy) increases
Assumption 2: when similar entities are closer, diversity (i.e., entropy) decreases

Claramunt added a spatial constraint of *intra-* and *extra*-distances creating a spatial diversity measure:

$$H_{Cl} = -\sum_{i=1}^{n} \frac{d_i^{int}}{d_i^{ext}} p_i \log_2 (p_i) \qquad (5)$$

where d_i^{int} is distance among the same class and d_i^{ext} is distance between classes. In Claramunt's equation, the $\frac{d_i^{int}}{d_i^{ext}}$ acts as a discriminant-ratio to assess how entropy changes with Euclidean distance. Claramunt argued that the spatial diversity measure increases when the distance between the same class members increases, or entities from different classes are closer in space. For example, in the chessboard case (Fig. 1), landscape B contains *intra-distances* that are larger and evenly distributed relative to landscape A, in

which the classes, green and white, are spatially clustered. In landscape C and D, entropy values are similar although both landscapes have different cell arrangements, which is something that should be carefully evaluated when comparing larger datasets.

While the incorporation of space into the Shannon entropy equation has proven to be effective for describing landscape patterns, the terms "near" and "related" in TFL used by Claramunt (2005) were vaguely defined (Miller 2004). Wang and Zhao (2018) introduced a new form of spatial entropy, which uses a mixture of Cushman's ideas on the total edge length and Claramunt's distances between classes. The *proximity factor* that the authors use is proportional to the number of edges between the distinct classes as L_i and inversely proportional to d_i, which represents an average distance between different class centroids:

$$H_{WZ} = -\sum_{i=1}^{n} \frac{L_i}{d_i} p_i \log_2 (p_i) \qquad (6)$$

Wang and Zhao (2018) argued that adding a proximity factor to entropy provides information about the spatial arrangement of entities in geo-space. This spatial entropy (H_{WZ}) idea was built on the following two assumptions:

Assumption 1: entropy increases as the TE between classes increases, which is the opposite of Cushman's assumptions.
Assumption 2: entropy should increase as the distance between different class centroids decreases, which aligns well with Assumption 1 from Claramunt's (2005) *spatial entropy*.

These assumptions were tested in our example, and the calculated entropy was low for the most aggregated mosaic in Fig. 1 (Fig. 1a) and high for the most heterogeneous landscape in B. H_{WZ} was able to distinguish between patterns C and D, which lacked clarity using Claramunt's measure. H_{WZ} was also easier to calculate, as only centroid distances and total edge lengths were used, although the computational difficulty may change when applied to larger biological landscapes.

Changing scale

We fine-grained the existing 3×3 landscapes A-D to 4×4, 8×8, and 16×16 (Fig. 1 and Table 1) to compare how spatial entropies change with spatial resolution, keeping the configurations similar to the original A-D landscapes. Unlike the Claramunt and Wang and Zhao spatial measures, Cushman (2016) did not provide easily reproducible examples and a way to compare results across different spatial scales. Additionally, the number of unique cell arrangements generally increases rapidly with increasing dimensionality, which makes the calculations prone to human error unless one has modelling programs like QRULE or FRAGSTATS (Gardner 1999; McGarigal et al. 2012). Therefore, we decided to avoid miscalculating Cushman's measure and chose to continue comparing how spatial entropy changes with scale only for the other two metrics (i.e., H_{Cl}, H_{WZ}). Claramunt's spatial diversity (H_{Cl}) followed a general declining trend with increasing resolution, with values dropping by approximately 15% from 3×3 to 16×16 scales. Wang and Zhao's spatial entropy (H_{WZ}) for the maximally aggregated landscape was lowest among all the landscapes, with E, I, and M landscapes showing equal results with increasing resolution. The maximally dispersed "chessboard" example B, F, J, and N increased with increasing resolution with landscape B showing lowest spatial entropy compared to F, J, and N because of the small ratio of total edge length and centroid distance between classes. In general, landscape N had the highest spatial entropy (H_{WZ}) value because of the extremely large TE value of 480, something that should be carefully evaluated when assessing highly dispersed landscapes.

The two entropic variants had different assumptions and used different variables in the final equations and thus, landscapes A-H had higher values since Claramunt's spatial diversity increases when entities from different classes are closer in space. The opposite was detected for Wang and Zhao's metric (H_{WZ}), where A-H landscapes had overall smaller values for H_{WZ}, likely because the 3×3 and 4×4 landscapes are quite similar when it comes to cell interaction range. In other words, A-D and E–H landscapes only have small spatial neighborhoods, whereas "close" and "far" neighbors start to be more evident in 8×8 and 16×16 landscapes.

The different ways statistical versions of entropy have been used in landscape ecology highlight steps forward in understanding the organizing principles of nature. Other efforts have been made to detect spatial patterns with Wasserstein entropy (Zhao and Zhang 2019; Zhang et al. 2020), decomposing original Shannon entropy (Altieri et al. 2018); or applying Boltzmann entropy to lattices, surfaces, and point patterns (Cushman 2021). However, it remains to be seen which methodological innovation will become most operationally feasible (Frazier 2019) at multiple scales. A solid statistical-mechanical and mathematical foundation for the proposed measures also still needs to be developed. Although the computation of Boltzmann entropy for landscape mosaics and gradients and its thermodynamic interpretations still requires further evaluation, the method is promising for detecting spatial patterns at many ecologically relevant scales. Additionally, the entropy indices described above often rely on changing the original formula, thus they should be viewed more as calculational variants than as a thermodynamic entropy. That is not to say that the entropy indices are not useful. On the contrary, Eqs. 4, 5, and 6 allow spatial patterns of landscapes to be studied, especially where composition, as well as proximity are significant in the analysis. Regardless of the preferred method, the concept of entropy is capable of capturing configurational and compositional information based on specific research questions. To continue using entropy as a statistical tool to describe a landscape's spatial patterns, we must be able to account for the multiscale nature of ecosystems and therefore create a mechanism that acts as a common currency across scales.

A common currency: Kullback–Leibler divergence

Kullback–Leibler divergence (D_{KL}), also known as relative entropy, describes how different the probability distribution P is from the previous estimate Q (Kullback and Leibler 1951):

$$D_{KL}(P\|Q) = \sum_{x \in \chi} P(x) \log \frac{P(x)}{Q(x)} \qquad (7)$$

It should be noted that D_{KL} mathematically speaking is not a distance but a divergence as

D_{KL} has two essential properties: non-symmetry $D_{KL}(P \parallel Q) \neq D_{KL}(Q \parallel P)$ and non-negativity $D_{KL}(P \parallel Q) \geq 0$ (Eguchi and Copas 2006). In botanical studies, D_{KL} was used for partitioning biodiversity measures (alpha, beta, and gamma diversities) by decomposing Shannon entropy into hierarchical nested levels. More recently, D_{KL} has been used to rescale temporal and spatial precipitation data to examine how different variables are affected by changing time–space scales (Sohoulande et al. 2019). The authors found that higher correlations were persistent on larger scales, suggesting that statistical interpretations are often influenced by a strong scaling effect on certain variables. In marine research, D_{KL} has been utilized to find variability between phytoplankton populations at different spatial scales that represent different habitats and environmental gradients (Roselli et al. 2013).

The D_{KL} is also a useful 'shortcut' for calculating the mutual information $I(X,Y)$, between the random variables X and Y:

$$I(X;Y) = D_{KL}(P(X,Y) \parallel P(X), P(Y)) = H(X) + H(Y) - H(X,Y) \tag{8}$$

where D_{KL} is the Kullback–Leibler divergence between the joint distribution $P(X, Y)$ and the marginal distributions $P(X)$ and $P(Y)$. H is the Shannon entropy mentioned earlier. Both definitions express that $I(X,Y)$ is the information shared by X and Y, i.e. what goes beyond the X and Y distributions individually. In contrast to D_{KL}, the $I(X,Y)$ results are interpreted in an inverse fashion: zero values show that X and Y are independent, higher values indicate a greater relationship between the variables (Villaverde et al. 2014; Corso et al. 2020). The symmetric property of $I(X,Y)$ may be preferred by some ecologists as it is often easier to interpret. However, even though D_{KL} asymmetry is more difficult to grasp, it contains more information than $I(X,Y)$. Imagine a predominant grassland with some weeds in it evolving into a diverse forest with no species dominant. Then D_{KL} (forest∥grass) $<$ D_{KL} (grass∥forest) because the forest is more uniform in its distribution. The forward and backward evolution are not equally likely, and therefore the difference is direction dependent.

Even with the current evidence, D_{KL} has been largely overseen by many ecologists as a potential statistics tool, often overshadowed by more common but not necessarily more well-founded similarity indices such as Bray–Curtis (Ludovisi and Taticchi 2006). The Bray–Curtis index gained its popularity for simplicity reasons—it is useful for detecting the extent to which the sampling units share, or do not share, similar biological entities (e.g., species, molecules, genes etc.). However, Bray–Curtis dissimilarity is biased towards dominant, disproportionally abundant species, under-representing less abundant species in the sample (Chao et al. 2006). In community-based studies, rare species can represent an important function or trait, thus disregarding them as "analytical annoyances" may limit a full understanding of the system (Jousset et al. 2017). This is where the logarithm of low probabilities in the Kullback equation could potentially be useful, especially in situations where conditionally rare taxa are important. Another valuable aspect of Kullback–Leiber divergence is the summation of the log of the ratio of each individual's probability between the two samples. Whilst a commonly used Shannon entropy only shows how much unevenness is in the probability distribution, D_{KL} can provide information on which species are causing the unevenness in the sample. Therefore, we propose to use the unitless property of D_{KL} that can differentiate between the samples across scales and add a spatial component to normalize D_{KL}, which would account for the spatial configuration within the sample.

The Kullback Information Index

Scaling and identifying spatial patterns start with a hypothetical sampling design shown in Fig. 2. Each sampling unit has an associated probability distribution of features (e.g., molecules, species, cover types etc.). In our model the sample consists of a mixture of two components, V and W. In Fig. 2, yellow represents 90% V, green 90% W (with an epsilon to account for possibly not sampling to completion and to mathematically deal with log 0). We can relate all the features of the system using a Kullback–Leibler divergence (D_{KL}), which describes divergence of one probability distribution from another, in this way creating edges between different nodes in the network of all the samples (Fig. 2b, e). The higher the D_{KL}, the more dissimilar the two distributions are, whilst a D_{KL} of 0 represents distributions that are identical. The resulting Kullback matrix can be averaged by the number of cells in the landscape (Fig. 2b, e) and thus
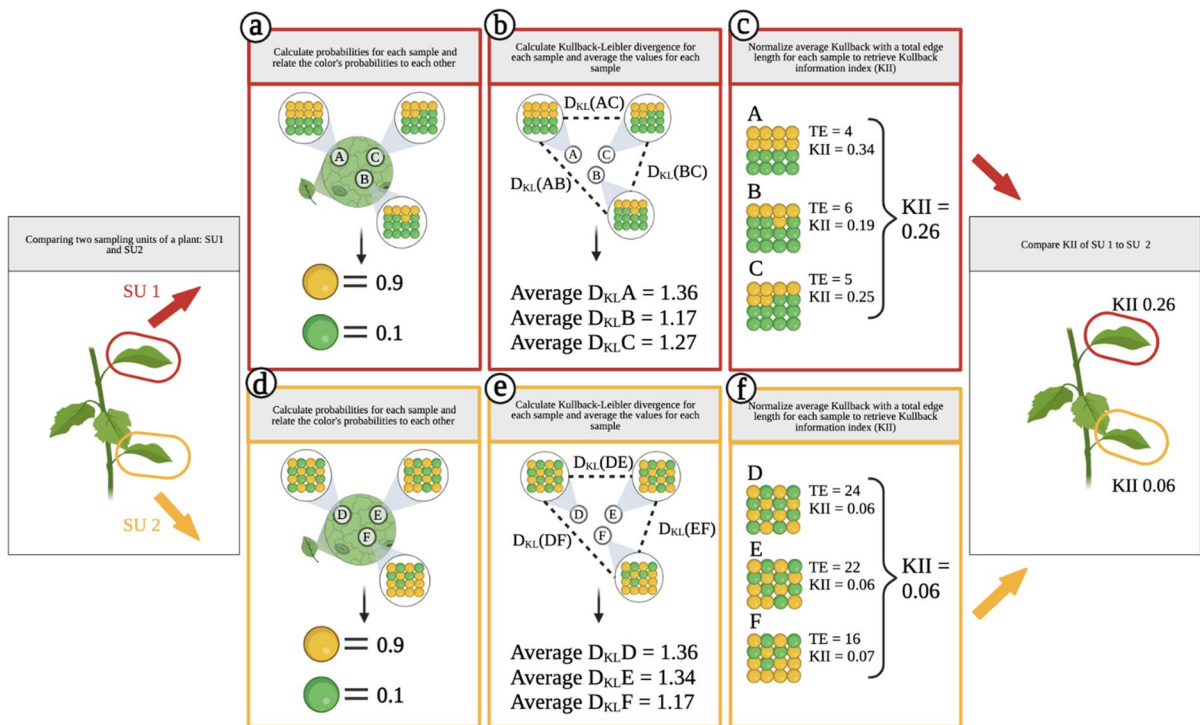
Fig. 2 A conceptual depiction of the framework that shows how to calculate average Kullback–Leibler divergence for a sample, and how to proceed to a Kullback Information Index (KII) that captures both the compositional and configurational organization of the landscape with a single number. In the top leaf (Sampling unit 1) the two types of cells appear in separate patches (yellow and green) in panel **a**. In **b** the Kullback–Leibler divergence D_{KL} is calculated for each sample and its average found for each location (A, B, C). In **c** each average is further divided by the edge length to yield the Kullback Information Index (KII), which thus is a property of the entire leaf. In the lower leaf (Sampling unit 2) the yellow and green components are thoroughly mixed in each sampling cell, yielding more or less the same averaged Kullback values **e** as in the upper leaf. However, the edge length is now much larger and the KII therefore noticeably lower **f**, clearly distinguishing the lower leaf from the upper. The KII can then be used to compare samples across scales by comparing to the maximum possible value of KII, which scales directly with the number of cells and inversely to the spatial normalization factor [i.e., Total Edge length (TE) in this example]

describes the average skewness of the community distribution within a sampling unit. To account for both probability distribution and spatial distribution, the average Kullback value needs to be normalized by a spatial component. In our example, we use total edge length (TE), which only accounts for nearest-neighbor interactions. In other words, TE provides 'spatial information' only when things are interacting locally. If other interactions are assumed or known for the observed landscape, other spatial normalization factors could be used to account for spatial organization (e.g., proximity). The averaged and spatially normalized Kullback measure is the Kullback Information Index (KII). The KII captures both compositional and configurational patterns in a landscape of any size

that can later be compared to other sampling scales (Fig. 2c, f). Higher KII depicts a highly aggregated system, whereas lower KII represents a highly dispersed system. By comparing systems with KII, one can detect the presence or absence of statistical relationships of spatial configurations across scales.

In summary, we propose calculating compositional information of each scale of interest, whether that would be a leaf or a polyp, or branch of a coral, and then investigate how each sampling unit is related across scales (Fig. 3a–f). The two numbers retrieved from the calculations provide two means of comparing samples. The average Kullback entropy describes the total distribution of types of entities in the whole system (Fig. 3b), whereas KII
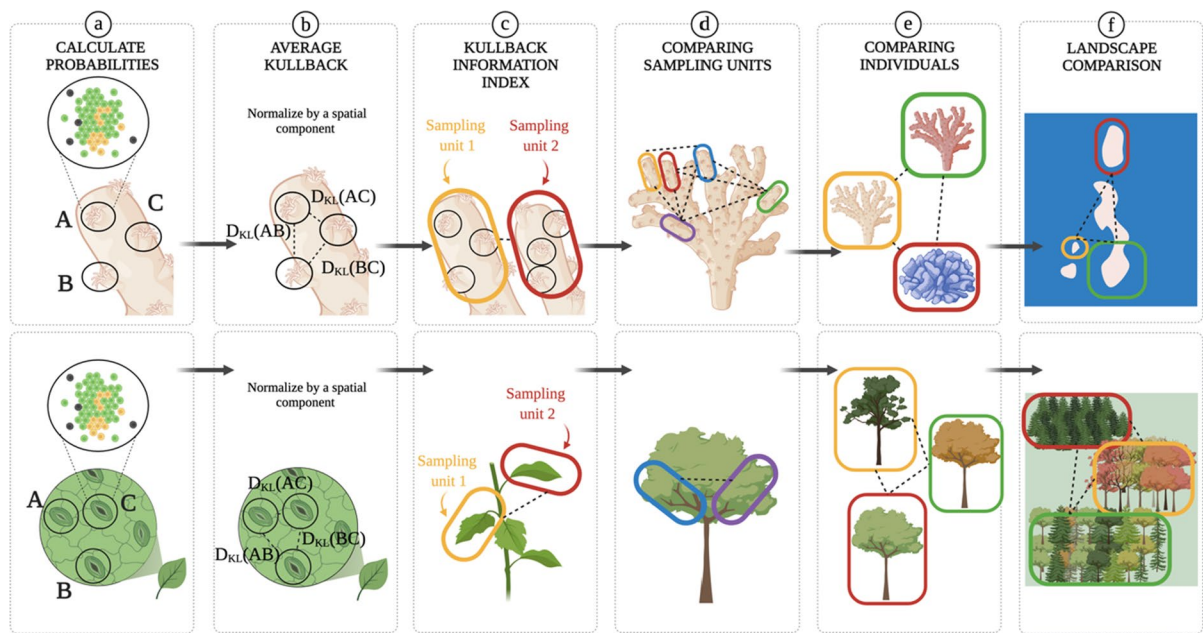
Fig. 3 A stepwise sketch of the proposal to study multi-scale entropy patterns of a coral reef (top panels) and forest ecosystem (bottom panels) using Kullback Information Index. The sequence is the following: **a** Calculate the probability distributions of the available components (green, yellow, black etc.) at different locations (A, B, C). **b** Calculate the Kullback–Leibler divergences and average them for each of these locations. **c** Normalize each of the averages with a spatial component (e.g., distance between samples or edge length) where the different components meet to get the Kullback Information Index (KII). **d** Compare these KII for different sampling locations. **e** Coarse grain further to compare different individuals and **f** still further to different landscapes

captures the spatial organization of that landscape with a single number that can be used across scales (Fig. 3c) by comparing values to the maximum possible value of KII, which scales directly with the number of sampling units and inversely to the spatial normalization factor (i.e., Total Edge length (TE) in this example).

Our framework aims to move from the traditional view of landscapes being large-scale observations, to any sub-space with multiple higher resolution sampling units. All the examples shown in Fig. 3 are landscapes on their own regardless of their size and scale. In that way our framework captures the statistical distribution of types of features (e.g., species, cover types, molecules, pixels) and spatial configuration of those features across hierarchical levels using the Kullback Information Index. This will open new avenues for understanding how "space" structures biological distributions and provides crucial information on the pattern-process relationship.

Discussion

Ecologists have a long history trying to understand spatial patterns in natural communities as these patterns often carry information on the ecological processes that shape them (Law et al. 2009; Turner 1989). The field of landscape ecology has successfully implemented entropy-like calculations to detect spatial heterogeneity. However, the variety of entropic measures used in the field show different results largely because they are built on different assumptions. For example, Claramunt's (2005) spatial entropy is purely based on Tobler's First Law of Geography, suggesting that all systems abide by these "rules". While this is true for some biological entities, these assumptions cannot be used for systems where organisms can actively pursue locomotion. Even for organisms that are sessile, these assumptions ignore the fact that other forces can be driving the spatial structure; for instance, the Janzen-Connell hypothesis

states that the fitness costs can be avoided if offspring do not cluster next to the parent (Janzen 1970; Connell 1971). In reef systems, some coral propagules are capable of long-distance dispersal due to currents, often blurring the notion that "near things are more related than distant things".

Furthermore, to calculate the number of unique arrangements within a landscape, Cushman (2016) used Total Edge (TE) metric, which measures the number of edges or boundaries shared between different classes in the system. The number of total edges within the landscape will depend on how entities are configured in space, and thus provide a Boltzmann-like entropic value describing that system. These ideas were incorporated in Wang and Zhao's (2018) spatial entropy, which used a combination of Cushman's Boltzmann entropy and Claramunt's distance calculations. In their formalism, the resulting entropy is equivalent to the degree of landscape heterogeneity where diversity increases when total edge length increases between different classes, and when different class centers are closer in space. Nevertheless, it should be noted that TE only accounts for nearest neighbor interactions, and unless functional arguments for assuming relevant interactions are considered, the entropy will not carry any physical meaning. Especially if future studies aim to relate patterns to thermodynamic processes, it is important to define what interactions matter, what conditions are held constant, and what is dynamic in the system. In statistical mechanics and physics where the concept of entropy originates, all interactions are quantified through the interaction energy as a function of distance between the interacting components. Not only can that interaction be either short or long ranged, but it can also be directional, and it is frequently not monotonic in its distance dependence. In other words, it could for example be a repulsion at short range and an attraction at longer range; this would create a distinct structure in the system's spatial appearance. It is likely that many structures in ecology are the result of similar non-simple distance dependencies.

There is no doubt that efforts to apply entropy in ecology have been fruitful. The probabilistic approach is attractive to many fields because both Shannon and Boltzmann entropies can statistically describe how systems are organized. Yet, the equations only consider the probabilities of entities in the

system without a reference to the entity type and are scale-dependent. To deal with the latter, an approach already used in physics is scaling with Tsallis entropy. Tsallis (1988) modified the original Shannon expression and used a power law instead of the logarithm in Eq. 2. Although originally applied in physics, Tsallis entropy has found its way to biological studies e.g., as a biomarker to find signs of dementia, or to improve gene regulatory network expression profiles (Lopes et al. 2011; Al-Nuaimi et al. 2015). However, Tsallis entropy does not provide configurational information that is necessary to examine how biological entities occupy physical or functional space. Thus, we introduce another way to scale systems accounting for spatial and probability distributions using the Kullback Information Index. The framework begins by estimating probabilities of each type of species of a sample and averaging the results over sampling modules within each sampling unit. The average Kullback is then normalized by a spatial component to account for spatial configuration within the system. For this step, different spatial normalizers such as total edge length, physical distance, or Wang and Zhao proximity can be used depending on the system and the interaction depth. The resulting KII will then scale linearly with the number of sampling units and inversely with the spatial normalizer.

Kullback–Leibler divergence has been previously used as a measure to account for information loss when coarse-graining (Shell 2008, 2012; Chaimovich and Shell 2010). The advantage of using D_{KL} in ecology over most other entropic measures, or dissimilarity indices, is that it compares the same species in two different distributions. This allows the construction of a geometry of probability distributions within the network (Shalizi 2007). Compared to Shannon entropy, which only demonstrates that something has changed, Kullback–Leibler divergence describes what has changed because it accounts for individual species probabilities. Since the measure is species sensitive, it could be used to study species change during stress-events or even compare perturbed and unperturbed systems (Dechant and Sasa 2020). Recently, Kullback–Leibler divergence has been used as an early-warning signal to detect complex diseases. Zhong et al. (2020) related reference samples taken from healthy individuals to a single 'unhealthy' sample and found that the single-sample Kullback–Leibler divergence method captures the tipping point before a

severe disease. Not only does this capture the impact of Kullback in medical sciences but also shows the potential applicability of D_{KL} in ecological studies where the early-warning signals could be useful to manage nonlinear ecosystem change.

It is worth mentioning that space is continuous, and the type of subdivision the researcher choses might lead to different results or the interpretation of those results. Even though the proposed framework would impose a certain division of biological units, the comparative power could be used to find data discontinuities within systems and provide additional information about how biology is structured in space or time. The design of this framework is advantageous because the Kullback Information Index shows the difference between the nodes in the network. In other words, this measurement offers different properties that can be used to study systems within and between hierarchical levels.

Conclusion

Identifying spatial patterns across a range of spatial and organizational scales is a complex task. Here, we show that building a theory that allows ecological data to be compared across scales is achievable. Our approach demonstrates how systems where spatial scales are nested can be described in a spatially explicit manner using a combination of probabilistic and spatial ecology approaches. We propose to average Kullback–Leibler divergence and normalize it with a spatial component to account for spatial organization of the system. The resulting Kullback Information Index accounts for not only compositional and configurational information of the ecological sample, but also is scalable. Future research could further extend the framework by describing the interaction terms between entities in the system, providing a deeper insight into biological relationships in space.

Data availability The data is available in a supplementary folder. The figures were created with BioRender.com.

Code availability Custom code is available as part of supplementary material.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication The authors consent the publication in Landscape Ecology.

References

Al-Nuaimi AH, Jammeh E, Sun L, Ifeachor E (2015) Tsallis entropy as a biomarker for detection of Alzheimer's disease. In 2015 37th Annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, pp 4166–4169

Altieri L, Cocchi D, Roli G (2018) A new approach to spatial entropy measures. Environ Ecol Stat 25(1):95–110

Boltzmann L (1872) Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen [Further studies on the thermal equilibrium of gas molecules]. Sitzungsberichte Akademie Der Wissenschaften 66:275–370

Chaimovich A, Shell MS (2010) Relative entropy as a universal metric for multiscale errors. Phys Rev E 81:060104

Chanda P, Costa E, Hu J, Sukumar S, Van Hemert J, Walia R (2020) Information theory in computational biology: where we stand today. Entropy 22(6):627

Chao A, Chazdon RL, Colwell RK, Shen T (2006) Abundance-based similarity indices and their estimations when there are unseen species in samples. Biometrics 62:361–371

Claramunt C (2005) A spatial form of diversity. Springer, Berlin, pp 218–231

Connell JH (1971) On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. Dyn Popul 298:312

Corso G, Ferreira GM, Lewinsohn TM (2020) Mutual information as a general measure of structure in interaction networks. Entropy 22(5):528

Cushman SA (2016) Calculating the configurational entropy of a landscape mosaic. Landsc Ecol 31(3):481–489

Cushman SA (2018) Calculation of configurational entropy in complex landscapes. Entropy 20(4):298

Cushman SA (2021) Generalizing Boltzmann configurational entropy to surfaces point patterns and landscape mosaics. Entropy 23(12):1616

Daly AJ, Baetens JM, De Baets B (2018) Ecological diversity: measuring the unmeasurable. Mathematics 6(7):119

Dechant A, Sasa SI (2020) Fluctuation–response inequality out of equilibrium. Proc Natl Acad Sci USA 117(12):6430–6436

Eguchi S, Copas J (2006) Interpreting kullback–leibler divergence with the neyman–pearson lemma. J Multivar Anal 97(9):2034–2040

Frazier AE (2019) Emerging trajectories for spatial pattern analysis in landscape ecology. Landsc Ecol 34(9):2073–2082

Gao P, Li Z (2019) Computation of the Boltzmann entropy of a landscape: a review and a generalization. Landsc Ecol 34(9):2183–2196

Gao P, Zhang H, Li Z (2017) A hierarchy-based solution to calculate the configurational entropy of landscape gradients. Landsc Ecol 32(6):1133–1146

Gao P, Zhang H, Wu Z (2021) Wasserstein metric-based Boltzmann entropy of a landscape mosaic: a clarification, correction, and evaluation of thermodynamic consistency. Landsc Ecol 36(3):815–827

Gardner RH (1999) RULE: map generation and a spatial analysis program. In: Klopatek JM, Gardner RH (eds) Landscape ecological analysis. Springer, New York, pp 280–303

Harte J (2011) Maximum entropy and ecology. Oxford University Publishing, Oxford

James PM, Fortin MJ (2013) Ecosystems and spatial patterns. In: Leemans R (ed) Ecological systems. Springer, New York, pp 101–124

Janzen DH (1970) Herbivores and the number of tree species in tropical forests. Am Nat 104(940):501–528

Jaynes ET (1957) Information theory and statistical mechanics. Phys Rev 106(4):620

Jaynes ET (1980) The minimum entropy production principle. Annu Rev Phys Chem 31:579–601

Jost L (2006) Entropy and diversity. Oikos 113:363–375

Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V, Hol WG (2017) Where less may be more: how the rare biosphere pulls ecosystems strings. ISME J 11(4):853–862

Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22(1):79–86

Law R, Illian J, Burslem DF, Gratzer G, Gunatilleke CVS, Gunatilleke IAUN (2009) Ecological information from spatial patterns of plants: insights from point process theory. J Ecol 97(4):616–628

Leibovici DG (2009) Defining spatial entropy from multivariate distributions of co-occurrences. In: International conference on spatial information theory. Springer, Berlin pp 392–404

Lesne A (2014) Shannon entropy: a rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics. Math Struct Comput Sci. https://doi.org/10.1017/S0960129512000783

Lopes FM, de Oliveira EA, Cesar RM (2011) Inference of gene regulatory networks from time series by Tsallis entropy. BMC Syst Biol 5(1):1–13

Ludovisi A, Taticchi MI (2006) Investigating beta diversity by Kullback–Leibler information measures. Ecol Model 192(1–2):299–313

McGarigal K, Cushman SA, Ene E (2012) FRAGSTATS v4: spatial pattern analysis program for categorical and continuous maps. Computer software program produced by the authors at the University of Massachusetts, Amherst

McQuarrie D (2000) Statistical mechanics. University Science Books, Sausalito

Miller HJ (2004) Tobler's first law and spatial analysis. Ann Assoc Am Geogr 94(2):284–289

Nowosad J, Gao P (2020) belg: a tool for calculating Boltzmann entropy of landscape gradients. Entropy 22(9):937

O'Connor MI, Pennell MW, Altermatt F, Matthews B, Melián CJ, Gonzalez A (2019) Principles of ecology revisited: integrating information and ecological theories for a more unified science. Front Ecol Evol 7:219

Rajaram R, Castellani B, Wilson AN (2017) Advancing Shannon entropy for measuring diversity in systems. Complexity 2017:1–10

Roach TNF (2020) Use and abuse of entropy in biology: a case for caliber. Entropy 22(12):1335

Roach TNF, Nulton J, Sibani P, Rohwer F, Salamon P (2017) Entropy in the tangled nature model of evolution. Entropy 19(5):192

Roach TNF, Nulton J, Sibani P, Rohwer F, Salamon P (2019) Emergent structure in a stochastic model of ecological evolution. Ecol Model 401:129–133

Roach TNF, Salamon P, Nulton J, Andresen B, Felts B, Haas A, Rohwer F (2018) Application of finite-time and control thermodynamics to biological processes at multiple scales. J Non-Equilib Thermodyn 43(3):193–210

Roselli L, Stanca E, Ludovisi A, Durante G, Souza JSD, Dural M, Alp T, Bulent S, Gjoni V, Ghinis S, Basset A (2013) Multi-scale biodiversity patterns in phytoplankton from coastal lagoons: the Eastern Mediterranean. Trans Waters Bull 7(2):202–219

Shalizi C (2007) 36–754, Advanced probability II or almost none of the theory of stochastic processes

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(3):379–423

Shell MS (2008) The relative entropy is fundamental to multi-scale and inverse thermodynamic problems. J Chem Phys 129:144108

Shell MS (2012) Systematic coarse-graining of potential energy landscapes and dynamics in liquids. J Chem Phys 137:084503

Sherwin WB (2010) Entropy and information approaches to genetic diversity and its expression: genomic geography. Entropy 12(7):1765–1798

Sohoulande CD, Stone K, Singh VP (2019) Quantifying the probabilistic divergences related to time-space scales for inferences in water resource management. Agric Water Manag 217:282–291

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46:234–240

Tsallis C (1988) Possible generalization of Boltzmann-Gibbs statistics. J Stat Phys 52(1):479–487

Turner MG (1989) Landscape ecology: the effect of pattern on process. Annu Rev Ecol Syst 20(1):171–197

Ulanowicz RE, Jørgensen SE, Fath BD (2006) Exergy, information and aggradation: an ecosystems reconciliation. Ecol Model 198:520–524

Villaverde AF, Ross J, Morán F, Banga JR (2014) MIDER: network inference with mutual information distance and entropy reduction. PLoS ONE 9(5):e96732

Vranken I, Baudry J, Aubinet M, Visser M, Bogaert J (2015) A review on the use of entropy in landscape ecology: heterogeneity, unpredictability, scale dependence and their links with thermodynamics. Landsc Ecol 30(1):51–65

Wang C, Zhao H (2018) Spatial heterogeneity analysis: introducing a new form of spatial entropy. Entropy 20(6):398

Wu J, Jelinski DE, Luck M, Tueller PT (2000) Multiscale analysis of landscape heterogeneity: scale variance and pattern metrics. Geogr Inf Sci 6(1):6–19

Zhang H, Wu ZW (2020) A head/tail breaks-based method for efficiently estimating the absolute Boltzmann entropy of numerical raster data. ISPRS Int J Geo Inf 9:103

Zhang H, Wu Z, Lan T, Chen Y, Gao P (2020) Calculating the Wasserstein metric-based Boltzmann entropy of a landscape mosaic. Entropy 22(4):381

Zhao Y, Zhang X (2019) Calculating spatial configurational entropy of a landscape mosaic based on the Wasserstein metric. Landsc Ecol 34(8):1849–1858

Zhong J, Liu R, Chen P (2020) Identifying critical state of complex diseases by single-sample Kullback–Leibler divergence. BMC Genomics 21(1):1–15

APPENDIX

Entropy is a statistical measure of uncertainty. In the realm of physics, it is often referred to as **thermodynamic entropy**, which describes uncertainty about the configuration of particles (microstates) within the specified system (a macrostate), resulting in the Boltzmann entropy

$$S = k \log W \tag{A1}$$

as it is chiseled into Boltzmann's tombstone in Vienna. The constant k defines the units you want for S. In the international SI system, S has units J/K in which case k is k_B, now known as Boltzmann's constant (1.389 x 10-23 J/K). W is the number of possible microstates corresponding to the specified macrostate. In most situations that means the number of microstates having the same energy as well as satisfying other possible other constraints.

In information theory entropy is calculated as the statistical average of the surprisal of each microstate, $\log p_i$, i.e., how unusual a particular microstate is:

$$H = -kN \sum_{i=1}^{n} p_i \log p_i \tag{A2}$$

This is known as Shannon entropy (sometimes referred to as **information entropy**). For large systems for which the Stirling approximation is valid, these two expressions are equivalent, as we will show here.

For a system consisting of N particles / microstates, grouped into M different macrostates the number of possible distinct manifestations is

$$W = \frac{N!}{\prod_{i=1}^{M} n_i!} \tag{A3}$$

where n_i is the number of microstates in macrostate i. For sufficiently large groups (which in reality is an amazingly small number: ~3), the factorial is well represented by Stirling's approximation,

$$\ln(K!) \approx K \ln(K) - K \tag{A4}$$

Using this approximation for W in *Equation A3* and entering the result into *Equation A1* above, we find that

$$\ln(W) = -\sum_{i=1}^{M} n_i \ln\left(\frac{n_i}{N}\right) \tag{A5}$$

$$= -N \sum_{i=1}^{M} p_i \ln(p_i) \tag{A6}$$

where $p_i = n_i/N$ is thus the probability of being in macrostate i.

QED, Boltzmann and Shannon entropies are the same. In both cases the constant k secures the desired unit for S in combination with the chosen logarithm (natural ln, decimal log, base-2 log, etc.). Please note that while the summation involves only intensive quantities (p_i, all in the interval [0;1] and independent of system size), the total number of microstates, N, in front ensures extensivity of S.

Supplementary material for
LAND-D-21-00313

Multi-scale spatial ecology analyses: a Kullback information approach

Gintarė HUCKEBA[1], Ty N.F. ROACH[2], Bjarne ANDRESEN[3]

[1] Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, The Netherlands
[2] Hawaiʻi Institute of Marine Biology, University of Hawaiʻi at Mānoa, Kāneʻohe, United States
[3] Niels Bohr Institute, University of Copenhagen, Denmark

A simple RStudio code (R version 4.1.0) was created to calculate spatial entropies of A-P landscapes from Figure 1. The results are presented in the main text Table 1.

The procedure starts with assigning each pixel coordinates in a separate data frame with an associated class (either green or white; in our case A or B, respectively). The code then aims to calculate Claramunt (2005) and Wang and Zhao (2018) spatial entropies.

Equations used in the R-script:

Claramunt (2005) added a spatial constraint of intra- and extra-distances denoting average distances within the same class (d_i^{int}) and between classes (d_i^{ext}), creating a spatial diversity measure:

$$H_{Cl} = -\sum_{i=1}^{n} \frac{d_i^{int}}{d_i^{ext}} p_i \log_2(p_i)$$

Wang and Zhao (2018) use the proximity factor that the authors use is proportional to the number of edges between the distinct classes as L_i, and inversely proportional to d_i which represents an average distance between different class centroids:

$$H_{WZ} = -\sum_{i=1}^{n} \frac{L_i}{d_i} p_i \log_2(p_i)$$

RStudio script

```
### CLARAMUNT'S SPATIAL ENTROPY ##
#A stand for green pixels, B stands for white pixels
df<- read.csv ("distances.csv",header = TRUE)
df= data.frame(df)

#Example Landscape A 3x3
#  x y class
#1 1 1    A
#2 1 2    A
#3 1 3    A
#4 2 1    A
#5 2 2    A
#6 2 3    B
#7 3 1    B
#8 3 2    B
#9 3 3    B

#Extract intra and extra distances from a Euclidean distance matrix

df %>%
  select(-class) %>%
  dist() %>%
  as.matrix() -> dist.m

as.matrix(dist.m)[grep("A", class), grep("A", class) ]

#Intra-distance for each class
intra <-sapply(LETTERS[1:2], function(let) as.matrix(dist.m)[grep(let, class), grep(let, class)
]
    )

#Extra-distance for each class
extra <- sapply(LETTERS[1:2], function(let) as.matrix(dist.m)[grepl(let, class), !grepl(let,
class) ]
    ) #this works as extra distance calculator

#Average Intra-distance for green class (i.e., A)
intra_green<-do.call(rbind.data.frame, intra[1])
intra_green<-mean(intra_green[intra_green!=0])

#Average Intra-distance for white class (i.e., A)
intra_white<-do.call(rbind.data.frame, intra[2])
intra_white<-mean(intra_white[intra_white!=0])

#Average Extra-distance for green and white classes (i.e., A and B)
extra_green<-mean(extra[extra!='A'])
extra_white<-mean(extra[extra!='B'])
```

```
#Probabilities for each class, for this example we are using Landscape A 3x3 probabilities:
p_g<- (5/9)
p_w<- (4/9)
p_g<- (p_g*log2(1/p_g))
p_w <- (p_w*log2(1/p_w))

#Claramunt's entropy ratios (intra/extra)
green_CL <- intra_green/extra_green
white_CL <- intra_white/extra_white

#Claramunt's spatial entropy
Claramunt <- sum(p_g*green_CL+p_w*white_CL)
print (Claramunt) #the result


######## WANG AND ZHAO SPATIAL ENTROPY #######
#Get centroid values for each class
centroids <- aggregate (. ~ class, df[, 1:2], mean)

# Split the centroid data frame into two observations
centroid_green <- centroids[1, ]
centroid_white <- centroids[2, ]

# Calculate and print their distance using the Euclidean Distance formula
centr_distance <- sqrt( (centroid_green$x - centroid_white$x)^2 + (centroid_green$y -
centroid_white$y)^2 )
centr_distance


#Calculate Wang and Zhao spatial entropy
#Entropy of that landscape
entropy <- p_g+p_w

WangZhao <- sum(10/centr_distance*entropy) #adjust according to each landscape's TE
print(WangZhao) #the result

#This code can be easily 'looped' for faster computing
```